Center for
Educator Compensation Reform
INFORMATION TECHNOLOGY CONSIDERATIONS

Christopher A. Thorn, University of Wisconsin-Madison
Robert Glover, University of Wisconsin-Madison
Jeffery G. Watson, University of Wisconsin-Madison

Editor, Cynthia D. Prince, Vanderbilt University.

The Center for Educator Compensation Reform (CECR) was awarded to Westat—in partnership with Learning Point Associates, Synergy Enterprises Inc., Vanderbilt University, and the University of Wisconsin—by the U.S. Department of Education in October 2006.

The primary purpose of CECR is to support the Teacher Incentive Fund (TIF) grantees with their implementation efforts through the provision of ongoing technical assistance and the development and dissemination of timely resources. CECR also is charged with raising national awareness of alternative and effective strategies for educator compensation through a newsletter, a Web-based clearinghouse, and other outreach activities.

This work was originally produced in whole or in part by the CECR with funds from the U.S. Department of Education under contract number ED-06-CO-0110. The content does not necessarily reflect the position or policy of CECR or the Department of Education, nor does mention or visual representation of trade names, commercial products, or organizations imply endorsement by CECR or the federal government.

INFORMATION TECHNOLOGY CONSIDERATIONS

There is no doubt about it; information technology (IT) in education is a complex field. Even though change in IT can be rapid, IT in education typically lags behind other fields such as business and medicine.[1] This lag provides education professionals some respite from having to be pioneers (an endeavor that is often fraught with risk), but it also creates more options when considering system design and development.

Traditional information systems in K-12 school districts focus on three different functions. First, districts need to meet their reporting requirements as identified by state and federal grantors and regulators. Second, districts need to manage a complex budgetary system and provide human resources support to a distributed set of facilities (schools). Finally, information systems have become increasingly integrated in the day-to-day management of school districts as school-level staff collect and enter attendance, discipline, and outcome data in local school offices or classrooms. This support for what is often identified as *transactional* data — data about changes in student characteristics — is often done both to distribute the workload of data entry (away from central or school offices) and is seen as vital to putting useful data in the hands of classroom-level staff.

Information systems have become increasingly integrated in the day-to-day management of school districts.

There are a number of challenges one must consider when exploring the links between various school information systems — particularly when the use of the data crosses traditional organizational and/or technical system boundaries. While there has been considerable success in creating ties between student information systems — such as enrollment, student grades, human resources, and free and reduced-price lunch programs — this success has not been universal.[2] The Schools Interoperability Framework (SIF) has been a leading organizer of education industry segments as well as of state, local, and federal agencies in its coordination of standard-setting for automated data exchange between systems.[i]

The emergence of the SIF standard for data exchange has allowed many districts to create links that had never before existed between information systems within districts. While this work has been quite useful, it has still addressed only a small portion of the data of interest for research and

[i] The SIF is a data-sharing specification originally developed to allow information systems within K-12 districts to exchange data without requiring wholesale replacement of existing systems. It includes both clear definitions for core data elements as well as secure methods for exchanging data. The Schools Interoperability Framework Association was founded to define the original standard and to provide a governance infrastructure for improving and expanding the standards-setting work. The SIF Association includes private software firms, state educational agencies, school districts and higher education institutions. The association has also expanded to include international members. See http://www.sifinfo.org for more information on the standard, the association, and its members.

program evaluation purposes. The first years have been focused almost exclusively on operational needs, such as maintaining accurate local address information and addressing increasingly complex federal reporting requirements. The integration made possible by data exchange standards has increased the availability of data, but it has also exposed quality problems when data are used for purposes other than those for which they were intended.

Incentive systems focused on increasing the productivity of schools (as measured by growth in student learning) must use technologies and social practices sufficient to track all critical elements associated with the production of this learning. A system focused on productivity must track inputs, such as funds expended for facilities and classroom resources, instructional staff members' characteristics, and curricula.[3] The system must also track the accumulation of knowledge in the form of professional development (PD) of the adults in the building — including amount and quality of the PD. There must also be a system of fairly recording the implementation of instruction. This should address the fidelity of the implementation of the curricula as well as core opportunity to learn information about the students in the room (e.g., who was pulled out, English language learner status). Finally, a productivity system must also include multiple measures of assessment for both the adults and the students. There should be an evaluation of teacher practice and application of content and pedagogical knowledge that feeds back to evaluate the delivery of PD and hiring practices. There should also be a robust set of measures of student learning that include both high-stakes and well-aligned diagnostic measures that align with clear, standards-based educational goals.

The problem with existing systems is that most compliance reporting does not include the correct data or does not require the appropriate level of disaggregation to evaluate productivity.

The problem with existing systems is that most compliance reporting does not include the correct data or does not require the appropriate level of disaggregation to evaluate productivity. One of the primary technical concerns in creating a viable pay-for-performance system is the significant existing investment in student information and personnel management solutions that districts have already made. These tools are deployed in all school districts of any size and manage the core administrative tasks of school districts. Unfortunately, most of these systems were not designed to manage student learning. Instead, they were designed to manage objects and activities such as classrooms, student-teacher ratios, and payroll. There are fundamental problems in the architecture of the technical systems that put the design focus of the system on objects and activities other than learning. This mismatch between system design goals and current user needs is a core problem in many districts.

In order to alleviate these problems, districts need to ask themselves several important questions about the systems they already have in place and about what they expect their new system to accomplish. To guide districts through this process, we have arranged these questions under four critical steps:

Step 1: Consider which people need to be involved in making decisions about what the system should look like and be able to do.

A. Are those using the system for decision support an integral part of the design process?
B. Are the groups that are subject to consequences provided with the time and resources needed to make a meaningful contribution to the system?
C. Do senior leaders make their commitment to the success of the project visible to all stakeholders?

Step 2: Determine what the school or district requires that the data system be able to do.

A. Does the system include performance ratings of outcomes and processes?
B. Linking student data to compensation raises data quality concerns. Is there an explicit acknowledgment by both leadership and those exposed to the incentive system of the connection between data-gathering mechanisms that collect high-quality data and their impact on the quality of analysis?
C. What time scales and units of analysis are important to address the analytical requirements of the model?
D. What is system integration and why is it important?
E. Who needs access and from where?

Step 3: Consider what the school or district already has in place and whether any of that is usable.

A. How can information systems address decision support needs?
B. Which source systems are relevant for meeting the decision support needs of those attempting to implement pay reform?
C. Why take an inventory of existing systems?

Step 4: Be aware of potential pitfalls that the school or district needs to pay attention to.

A. Are there particular areas of concern one should watch when choosing which features to emphasize when deploying data warehouse and other decision support systems?
B. Are there design and build pitfalls that one should monitor as part of the deployment of data warehouse and related decision support systems?
C. What should district leaders know about potential pitfalls in data warehouse design?

Step 1 Consider which people need to be involved in making decisions about what the system should look like and be able to do.

A. Are those using the system for decision support an integral part of the design process?

Many IT projects suffer by having only the leaders of district-level departments and relevant IT staff at the table when services and products for building-level use are designed and approved. The logic of this is all too easy. The people who pay for the project and are responsible for the overall results of the program all work at the district office. It is to those people (the business process leader/owner) that the IT staff members are responsible. It is quite easy to forget that the end users need to provide data, consume interim feedback, and are the people who will ultimately be judged by the results (along with their students).

B. Are the groups that are subject to consequences provided with the time and resources needed to make a meaningful contribution to the system?

This is a corollary to the question above. It is not only access to the process that needs to be guaranteed. Building-level leaders and classroom teachers also need the time and training to make a useful contribution. This is a classic case of *people don't know what they don't know*. System designers have very little knowledge of the ways in which building-level staff members use data for daily and weekly decisionmaking. Likewise, teachers and instructional leaders are usually unfamiliar with the details of formal state- or federal-level reporting or the complexities of value-added models. There will need to be a significant investment in professional development and dialogue for both sides to reach a rich understanding of system needs and design goals.

C. Do senior leaders make their commitment to the success of the project visible to all stakeholders?

Displays of high-level commitment of resources and political support are key success factors in any complex, costly project. Clear articulation of the goals for the project and the availability of the necessary infrastructure are probably the two best predictors of success. This is even more important in IT projects. IT projects are often very expensive and not well understood by non-technical staff members. The existence of high-cost projects with opaque goals will lead to active resistance, or at least a lack of faith that the project will be supported to completion.

Step 2 Determine what the school or district requires that the data system be able to do.

A. Does the system include performance ratings of outcomes and processes?

While there may be relative agreement on the analysis and presentation of test score outcomes and related value-added analysis, the collection and analysis of teacher performance ratings has left many districts struggling to address issues of equity and fairness. There are key questions about the appropriate level of correlation between performance ratings and value-added measures. From a system design point of view, there are also important design considerations around the collection and presentation of such data. There are concerns about the use of such data for high-stakes decisions (teacher pay or sanctions) if the validity of the data is in question. The results are also far more accessible to all who need to make use of them if they are collected in a district information system rather than in a notebook in a building leader's office. The inclusion of performance rating metrics in district IT systems will introduce a level of scrutiny that is likely to cause some concern at all levels in the district.

There needs to be a significant investment in professional development and dialogue for both sides—system designers and teachers and instructional leaders—to reach a rich understanding of system needs and design goals.

B. Linking student data to compensation raises data quality concerns. Is there an explicit acknowledgment by both leadership and those exposed to the incentive system of the connection between data-gathering mechanisms that collect high-quality data and their impact on the quality of analysis?

This is a core question that will likely be a good predictor of the successful pay-for-performance system. The answer to this question is likely to change as the data in question move from being merely routine compliance data (counts) to high-stakes information about productivity. Information quality is increasingly seen as one of the most important aspects of information system design within large institutions. Unfortunately, this focus on quality may be limited to an examination of the data by technical system administrators. Larry English and other researchers encourage developers to step back and acknowledge data users as the arbiters of information quality.[4] In a school setting, there is a wide range of information users, as discussed in greater detail later in this chapter. English argues that,

> Quality is a *customer* determination, not an engineer's determination, not a marketing determination or a general management determination… based upon a *customer's* actual experience with the product or service, measured against his or her *requirements* — stated or unstated, conscious or merely sensed, technically operational or

> entirely subjective. Because knowledge-workers require information to perform their work, the definition of information quality — (consistently meeting all knowledge-worker and end-customer expectations through information and information services) — focuses on them.... Therefore, a true quality method must focus on information producers as important people in the information value chain.... Many information quality problems today come from data that meet only one narrow set of customer needs (*my* department).[5]

School- and district-level report cards might meet the quality requirements of some audiences, but like individual student report cards, they provide a rather simplistic aggregation of what is a rich and nuanced process. The perusal of a student portfolio or a teacher's grade book and lesson planner provides a much better representation of the scope of data that could be incorporated into a richer decision support system. The point made by English that should be taken up by information workers and community members outside district IT departments is that it is data users who are the ultimate arbiters of data quality and data's usefulness in making decisions.

C. What time scales and units of analysis are important to address the analytical requirements of the model?

Another technical concern is the need to deploy systems that can respond to the various reporting and analytical needs of users at different levels of the organization. These users will need data at various levels of aggregation and for different periods of time. Teachers need very timely access to data that will help them work with individual students (and groups of students) to achieve learning goals across a specific span of instructional time — course unit, marking period, semester, etc. The relative autonomy of schools and teachers makes service provision in this area particularly difficult. Individual teachers have great latitude in determining what information they need to manage and report on student learning. It is nearly impossible to do a needs analysis if the units of analysis and time scales are different across individual classrooms.

The inability to anticipate the needs of teachers systematically in a heterogeneous environment is probably the most significant barrier to pushing decision support services down past the school and grade levels. While one can envision making integrated grade books and lesson planners available to individual teachers, these resources contain data that are idiosyncratic and pertain only to those students in that classroom at that time. This is a particular challenge for pay-for-performance systems, if one of the core goals is to measure classroom productivity. While one might be able to generate productivity estimates for individual classrooms or teachers, the data necessary to understand the practices that led to the observed growth are often difficult to collect and analyze. While the primary intervention of the incentive system is to reward teachers for outstanding performance, the logic of change underlying the policy is to show the path for others to achieve the same outcomes. The only way to do this systematically is to be able to show the path (e.g., resources available, instructional practices) to the desired state.

One of the most pressing concerns for IT system builders in light of the requirements of *No Child Left Behind (NCLB)* legislation is creating the capacity to demonstrate improvement in student learning outcomes using longitudinal test data. The problem with this focus on the annual test outcomes is that school-level staff members assert that this neglects the information needs of local decision makers as they address the problem of improving student learning. If one looks at the informational needs of teachers or building-level leaders, their need for useful information is focused on very frequent, low-stakes measures. Annual tests are so thoroughly divorced from the decisions that a teacher needs to make in the classroom on a daily basis that there seems to be an almost complete disconnect in the minds of many teachers. There are also fairly straightforward technical arguments against using standardized test scores for diagnosis of student needs.[6] These tests, by definition, provide a limited picture of the full range of desired learning outcomes. They cannot possibly measure the full range of student knowledge (or lack thereof).

One of the most pressing concerns for IT system builders in light of the requirements of **No Child Left Behind** *legislation is creating the capacity to demonstrate improvement in student learning outcomes using longitudinal test data.*

While annual standardized tests can provide important feedback about the general state of an educational system (as measured by the relative breadth of the test instrument), they do not provide the wealth of information a highly complex educational system needs for making most of its operational decisions. The following tables represent a radical simplification of the various roles and information flows in play within a school district and a representation of the informational needs, inputs, and outputs of individuals in those roles.

These tables serve at least two different purposes. First, they allow one to consider the information flows related to particular roles within the educational system and between the different roles. Second, they provide a checklist for assessing the capacity of existing systems to meet the informational needs of people in each role. While both needs are important, the second has become increasingly pressing in the face of *NCLB* requirements that districts and states meet new demands for detailed reporting requirements that require access to disaggregated data that are likely to stress or overload existing reporting systems.

Table 1. Information Flows Into and Out of Educational Information Systems

Role	Provide Data to Educational Information Systems	Get Results Out of Educational Information Systems
Students	X	X
Parents	X	X
Teachers	X	X
Instructional Support	X	X
Special Services	X	X
School Administrators	X	X
District Administrators	X	X
Community Members		X

The first column in Table 1 provides a list of system roles that are meant to represent different stakeholders across an educational system. One could come up with a number of additional categories, but for simplicity's sake, *Instructional Support* can be thought of as including teachers' aides, learning resource specialists, or other personnel who support regular education activities. *Special Services* represents staff providing English as a second language, special education, and therapeutic services. The role of *School Administrators* extends beyond the principal and assistant principal to others with managerial or supervisory responsibilities. By the same token, *District Administrators* can be thought of as including program managers across the various areas within a district office. The remaining two columns in Table 1 address whether any particular group provides or gets out (extracts) the data for a particular task or tool. The point of this table is that almost everyone in an educational system is both a provider and recipient of district data — either at the aggregate or individual level.

Table 2. Information Flows Up and Down by Organizational Level

Role	Provide data for use at lower level of organization	Provide data for use at higher level of organization
Students		X
Parents		X
Teachers	X	X
Instructional Support	X	X
Special Services	X	X
School Administrators	X	X
District Administrators	X	X
Community Members		

Table 2 makes a different and more subtle point. The second and third columns indicate whether or not data entered/generated are provided to a higher level of the organization (one or more levels up in the hierarchy) or to a lower level. While users at almost all levels of the system provide and receive data, there is also an important hierarchical element to the flow of data. Data are often collected at one level of the organization for use at another. While this may sound simple and obvious, this separation between the providers and users of data lies at the head of many data quality problems. Rules for collection and use at one level may not be clearly communicated across these organizational levels.

The question here is whether the tools in use at any particular level of the educational system support the creation/definition, management, analysis, and use of local data at the same level (or for sharing between adjacent levels). The core problem is to design technical systems and social processes that communicate the potential uses of data and impact of problems with validity and reliability.

D. What is system integration and why is it important?

Integration is important because the source systems that a district typically uses do not easily 'talk' to each other and therefore effort must be spent to try to integrate the data that these systems manage.

Integration work involves pulling data from source systems and joining those data through some common identifier. For example, districts typically will join student enrollment data for their SIS with assessment data from their assessment systems by joining through a student identification number. Integration is important because the source systems that a district typically uses do not easily 'talk' to each other and therefore effort must be spent to try to integrate the data that these systems manage. Integration in general is an issue that is affecting many districts, and some systems may be easier to integrate than others. The basis for many integration problems is the fact that source applications often have unique, sometimes proprietary, data definitions and structures. For example, one system may handle individual names with three fields (Last Name, First Name, Middle Initial) whereas another system might have one field (Name). In this case, matching records from both systems on 'name' becomes difficult due to inconsistencies in the data definitions.

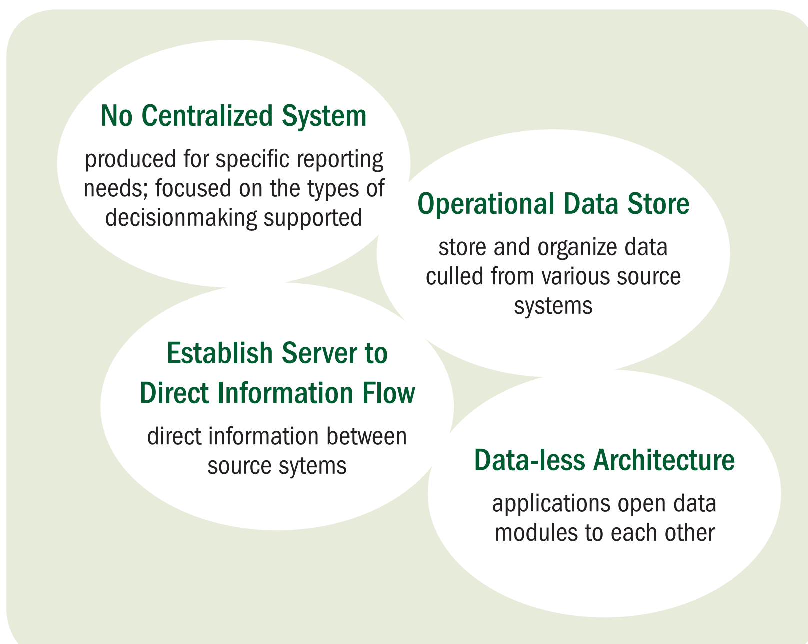
In short, what is required when an organization engages in IT development and integration work is a data dictionary that specifies and defines the data elements from the systems in use by a district. When data dictionaries exist as formal documents, they are usually thought of as code books that map field names into more descriptive language, and the field is described as either numeric, string, or categorical. If categorical, the categories used are usually listed and defined. Often districts do not maintain code books for their data files. In this case, development efforts by non-district staff can be cumbersome and more costly. Furthermore, it is difficult to assess whether or not a system can support a particular type of decision support if administrators and programmers are unclear about the details of their source systems.

The Schools Interoperability Framework Association (SIFA) has grown, in part, because vendors and stakeholders agree that systems should adopt a basic data model to reduce development time and improve interoperability between systems. SIFA has focused on providing a standard data model and a testing harness for software agents designed to push and pull data between each other via an integration server. SIFA has also recently released a web services model for vendors looking to develop web-based applications.

Many options are available to districts for integrating data across source systems, and it is beyond the scope of this article to try to compare and contrast these approaches. However, we can describe four common approaches. The first approach involves no centralized system to manage integration. Usually, integration is done in an ad-hoc fashion and will result in multiple data files that have been merged through various methods. These files are usually produced for specific reporting needs, such as the production of school-wide averages of test scores, and are focused on the types of decisionmaking that may be supported.

Another approach involves using an operational data store (ODS) to store and organize data that are culled from various source systems.[7] These data files are often created by running extract, transform, and load routines on a somewhat regular schedule. They may be held in ODS until needed, or they may be cleaned and loaded up to a data warehouse for archival and analytic purposes. Another approach to integration involves establishing a server to direct information flow back and forth between source systems as needed. Information is published and subscribed to through the implementation of software agents that communicate between source systems and the integration server.

Figure 1. Four Approaches to Integrating Data Across Source Systems



Likewise, a district may use a data-less architecture that utilizes an approach where applications open their data models to each other.[8]

Instead of sending data up to an ODS or an integration server, applications call data as needed. This data-less approach is generally untested in education, though other fields that have begun moving to a services-oriented architecture are more likely to attempt these types of projects.

E. Who needs access and from where?

One of the concerns that arises during system design is the difficulty of judging who needs access to what information resources. The use of paper reports based on data that only change annually or semi-annually is probably not a good predictor of use. While the data on teacher, school, grade, or classroom productivity might be the result of an annual analysis, the rising stakes of that payout (and any related sanctions for poor or declining performance) will encourage both leaders and teachers to pay more attention to the interim metrics that might provide some prediction of how the high-stakes results will play out.

What this means for system planners is that a wide range of users may begin demanding more access to interim data. Some of these needs might be addressed by providing teachers with online grade books and lesson planners to plot student progress systematically and bring together interim tests in one place. This would also support team- and building-level discussion about students' progress across subjects and grades. These sorts of systems are being deployed in many school districts. They require substantial support resources, but they also provide an opportunity to deliver annual high-stakes data back to teachers and building leaders for them to use in the local context. The availability of local systems containing local data is an important component in making high-stakes data real and relevant at the building level.

Step 3 Consider what the school or district already has in place and whether any of that is usable.

A. How can information systems address decision support needs?

Information systems focused on teacher and school productivity fall under the heading of transactional systems (data gathering), compliance reporting, or decision support (presentation of actionable results). Several challenging areas are associated with implementing decision support in K-12 districts and schools. The first challenge is to identify the informational needs of users in different roles across the educational enterprise. The second critical task is to explore these scenarios to identify the decision support needs that are left unmet by currently fielded systems.

Kesner defines decision support as "an IT-enable[d] system that facilitates the integration of critical enterprise information so that management may employ that information to inform planning and decision making."[9] One of the core concepts of system design is the notion of *granularity*. Data can be thought of in terms of the size of the most basic building blocks or units of analysis of the system. The granularity (or grain size) refers to the smallest observable unit for a given dimension of interest. At the elementary school level, for example, many accountability systems are unable to link individual students to teachers by subject area taught. The data systems simply do not track that level of data. In this case, the grain size for the student-teacher link would be at the classroom level.

An additional distinction that Kesner and others cite is that the *temporal granularity* (relevant time scale for a particular decision) varies by an individual or team's position in the hierarchy of an organization.[10] For example, a district curriculum team charged with deploying a new reading program to primary grades might combine grade-level historical data on test outcomes in the subject with current-year tests as well as end-of-unit tests in a predictive model to examine current learning growth as compared to historical patterns. A building-level team might, on the other hand, look at weekly reading diagnostics for the current grading period to make decisions about ability groupings for the next week — with no connection to historical data. Both would be examples of data used for decision support, but the data used and the systems deployed to support the work would be different. The important distinction here is that a district and a school require different levels of temporal granularity for information to be useful to them.

The granularity (or grain size) refers to the smallest observable unit for a given dimension of interest.

B. Which source systems are relevant for meeting the decision support needs of those attempting to implement pay reform?[ii]

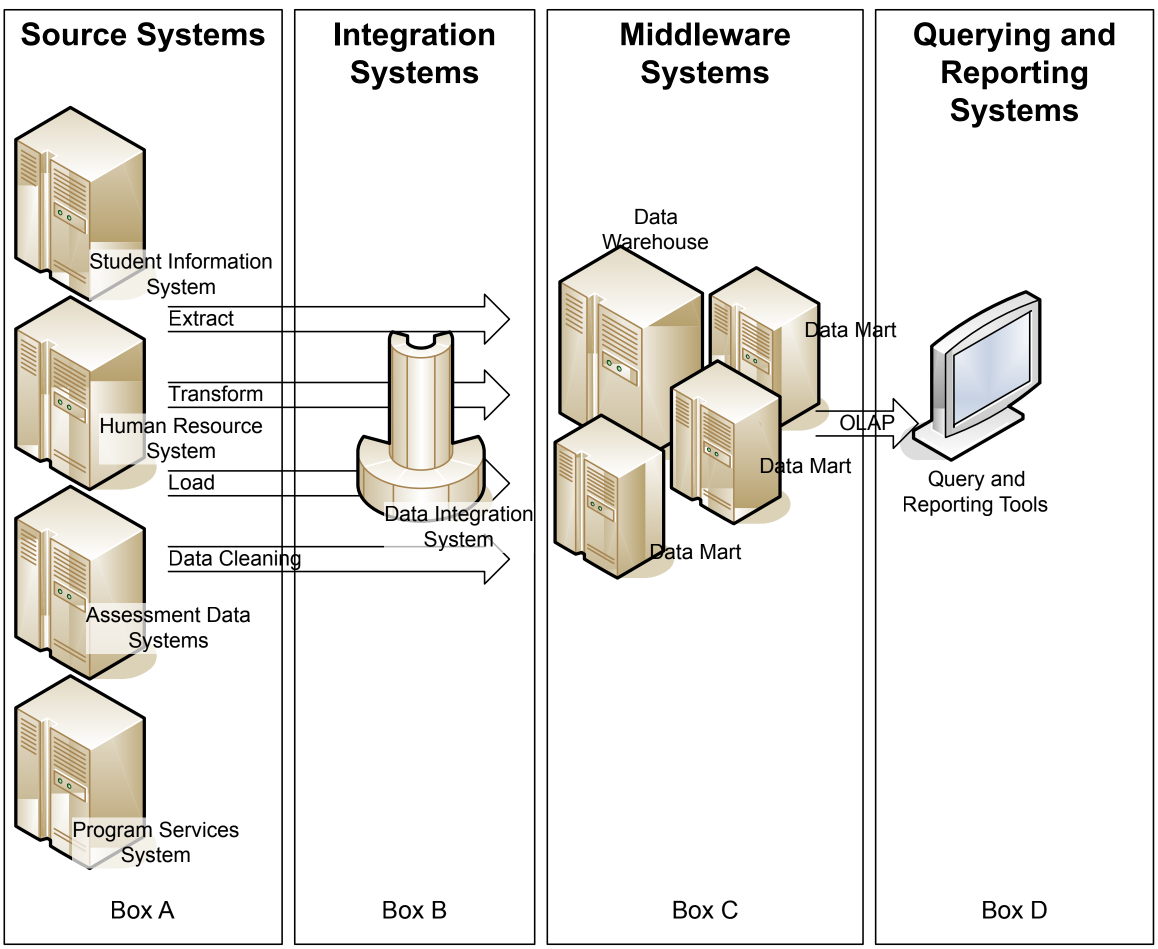
While they might not be described as such, the role of most information reporting systems is to support human decision makers. The generic term for such services is decision support systems. These systems are designed to aggregate data and implement analytical models that provide a framework for decisionmaking at each level of the organization. If we limit our discussion simply to those systems that support the management of students and faculty and analysis of outcome data and those other systems that might be used to track production processes, the universe of IT in education is distilled to just those systems that manage process and outcome data rather than those that seek to enhance learning environments (see Figure 2, box A). Typically, a district will use one student information system (SIS) to track student enrollment. While there are differences in how SIS might be implemented, it is often designed to manage the workflow around assigning students to a school and a set of courses.[11] The SIS will also be used to assign teachers to courses

[ii] The term *source systems* refers to the information systems and associated practices used to collect and manage data across the educational system. There are many source systems for an accountability system — including assessment, human resources, certification, transportation, food service, and English language learner (ELL) services. Many of these data systems have online systems for both entering and viewing data. They are often supported by reporting systems as well as by manual collection and verification of missing or suspect data.

(and therefore to students), and it often manages data about attendance, grades, demographics, health information, and discipline referrals.[12]

Other source systems that school districts rely upon include payroll, assessment, and supplemental services. Perhaps second in size and complexity to the SIS is the system that a given district uses to manage its payroll or human resource (HR) functions. The HR system usually tracks teachers' data within the context of maintaining certification and accreditation and managing payroll functions, and professional development in some cases. An HR system is typically not used to manage course assignments, but will often track school assignments. Assessment systems represent another large system for districts, though they can be quite variable depending on whether or not the district uses a state system or has a periodic assessment system, and the nature and quality of achievement data files that come from state or test vendor systems. Other types of systems that districts use include systems to track and manage professional development activities, supplemental services (e.g., tutoring, after-school programming), nutritional services, and identification card systems.

Figure 2. Typical Information Systems in K-12 Districts



C. Why take an inventory of existing systems?

In her work on community colleges, Petrides identifies the relative cost of information system purchases as second only to capital improvements in size.[13] She also points out that while capital improvements can be amortized over decades, information systems rarely last a decade and are often obsolete after three or four years. Given the costs involved and the related risks, it is not surprising that K-12 school districts are reluctant to give up systems that have already been fielded, but from which they have not yet wrung the expected value. More important, changes in educational support systems often have system-wide implications for school-level staff. Reforms that acknowledge the needs and capabilities of the people who will operate and use these systems are much more likely to succeed. A thorough inventory of existing systems and their information flows will provide a much better picture of unmet needs, unused capacity, and bottlenecks.

There are examples of the importance of performing system inventories from other areas of public policy. One consulting group that builds systems for public agencies that provide health care and economic aid has published an interesting set of white papers that acknowledge that the substantial investments that states and municipalities have made in large-scale, older technologies precludes any rapid movement to new tools.[14] Hamblin agues however, that even these older systems can be made better by improving the quality of the data being put into them, and that the customer experience can be enhanced by providing better and more timely feedback that moves the process forward by guiding the recipient through the application process.[15] There could be a number of parallels to the classroom-level tools. One example would be in support of demonstrating opportunity to learn (OTL) — something that is often a concern in high-stakes systems but that is rarely adequately addressed. Other uses might include providing individualized progress reports to parents and students as students move toward mastery in various learning goals, or non-technical explanations of challenging curricula for students to share at home.

One negative example emerges from research on patient information systems used by hospitals.[16] The author cites a study by Deloitte Touche of the "100 Most Wired Hospitals/Health Networks," in which two-thirds of the sites listed did not provide anything that would "save [a visitor] the trouble of calling, visiting, or sending mail to the organization or to its physicians." What this means is that the majority of sites were designed to provide information as a one-way flow and not to support interaction. This would very likely describe most K-12 school districts in the United States, and almost all schools. Of the sites that did provide some level of interaction, most were not associated with getting care. Rather, sending flowers, cards, or gifts to patients; applying for jobs; and taking a tour of the facility were more likely features. Of the most highly interactive sites, there were several that allowed physicians to gain secure access to patient information. There were also sites that allowed patients to create health records for themselves or for their children, but not one of the secure patient sites provided a method for communicating directly with one's physician.

One might ask what this has to do with design of school information systems. There are some very strong parallels and a lesson that is quite daunting. HMOs, in particular, have a number of similarities with school districts. They often have geographic memberships that are complicated by other choice-based affiliations. Data are collected in great detail on the symptoms presented by each individual patient at the point where they receive primary care, but those data are usually stored locally in handwritten form. Only aggregate data are used to measure the output of the physician or care team. As in the case of public schools, patients who go to a care center other than their primary site are only identified as belonging to the system, and very few data are available for the local caregiver to use to provide context for diagnosis. The patient is often the only source of contextual information — with the inaccuracies that implies.

As in the case of public schools, patients who go to a care center other than their primary site are only identified as belonging to the system, and very few data are available for the local caregiver to use to provide context for diagnosis.

The disturbing factor in the case described above is that HMOs have the profit motive working for them — a force that is not at work in public schooling. In fact, several health system developers have been fielding systems that would provide strong links among HMOs, physicians, and patients. Epic Systems espouses a patient-centric design philosophy that makes patient chart records the central point around which all online (and most offline) interactions are focused.[17] Many developers and hospitals have attempted to field integrated systems since the mid-1990s. It is only in the last few years that internet technologies have reached a level of maturity that secure, thin-client web technologies are ubiquitous.[iii] This enabling technology will very likely put these transaction cost-reducing patient-physician systems on the front burner for many health care organizations. The lower deployment costs have led to the widespread use of real time electronic record-keeping for office visits or other routine care. Patients are often encouraged to log into their individual records from home to communicate directly with a range of health care providers, as well as to access additional information resources. There is great potential for this model in the education sector.

[iii] *Thin-client* is a term used to refer to client-server applications in which the bulk of the processing happens on a central server and the client, or end user, needs only a browser to interact with the system. This means that software installation becomes much easier and that the resources become widely available because of this easy access.

Step 4 Be aware of potential pitfalls that the school or district needs to pay attention to.

A. Are there particular areas of concern one should watch when choosing which features to emphasize when deploying data warehouse and other decision support systems?

While most trade publications and vendors would urge districts and state agencies to build or buy data warehouse systems, there are studies from the private sector that suggest that one should be rather cautious. The starkest outcome of recent research is that there is a fairly wide range of warehouse project goals, and which goals are chosen will likely have a large impact on one's chances of success. The following list outlines the categories that emerged from a large set of case studies of large corporations that engaged in system development of knowledge management systems.[18] While commercial systems do not map directly on educational systems, there are enough similarities that this review is a helpful one. System development activities fell into four large categories:

1. **Create knowledge repositories –**
   a. External knowledge (e.g., competitive intelligence, market data, surveys);
   b. Structured internal knowledge (reports, marketing materials, techniques and methods); and
   c. Informal internal knowledge (discussion databases of 'know how' or 'lessons learned').
2. **Improve knowledge access through:**
   a. Technical expert referral;
   b. Expert networks used for staffing based on individual competencies; and
   c. Turn-key video conferencing to foster easy access to [geographically] distributed experts.
3. **Enhance the knowledge environment –**
   a. Change organizational norms and values related to knowledge in order to encourage knowledge use and knowledge sharing; and
   b. Customers may be asked to rate their provider's expertise.
4. **Manage knowledge as an asset –**
   a. Attempt to measure the contribution of knowledge to bottom line success.[19]

Knowledge repositories (Category 1) have their analogues in educational systems. Recent efforts to develop standards-based learning object collections or less ambitious curriculum repositories are examples of systems that attempt to present external knowledge in a useful form.[20] Internal systems that provide *real time* school performance feedback fall into the category of structured internal knowledge systems. The most difficult system to build in this category is the *know-how* collection that would provide detailed context information about how a particular policy was implemented or how a new curriculum was rolled out successfully.

Systems designed to improve access to knowledge (Category 2) are focused on using technology either to help users with a particular need to locate an expert and/or to provide technical support to overcome barriers to collaboration and communication. A system that links beginning teachers or administrators to a team of mentors who can provide timely, expert guidance on demand would be an example of this sort of system. Systems that track the details of implementation of successful and unsuccessful innovative programs would be another example.

Systems intended to enhance the knowledge environment (Category 3) may actually be the most likely result of the pressures of *NCLB* and pay-for-performance systems. Value-added analysis and its resulting improved accuracy in representing how much and under what conditions students are learning should provide a fair and accurate representation of what is and is not working within school districts. Districts that are able to use this information to improve local practices will be building ties between their reporting systems and classroom practices — a situation that is often not the case in systems that look only at annual attainment.

The other important aspect of this category (knowledge enhancement) is that systems are emerging that allow community members to provide feedback and ask questions of teachers and administrators. Some places have experimented with systems that allow students and parents to rate the feedback that they are getting from teachers. Districts that have decentralized budgetary and other authority to local schools are getting anecdotal feedback when schools refuse to buy back services that they do not value. The prevalence of free or low-cost collaboration tools, the growing speed of district computer networks, and the increasing computer literacy of school staff suggest that systems may emerge that reflect school feeder patterns or other groupings that do grow out of district-level efforts.

Assessing the value of knowledge as an asset (Category 4) remains as elusive as earlier attempts to determine the return of investment (ROI) on research expenditure.[21] Davenport and his colleagues did not find a single example of a system that had been successfully fielded.[22] Given the relative immaturity of decision support system development in educational settings, it is unlikely that one will see any attempt to quantify the ROI of knowledge management or decision support systems to student learning. It is not too early, however, to begin to examine how one might evaluate the systems described in Categories 1 to 3.

B. Are there design and build pitfalls that one should monitor as part of the deployment of data warehouse and related decision support systems?

One of the single most comprehensive descriptions of the common pitfalls of data warehouse deployment is available as part of an online course at the Operational Research Society. Table 3 provides a synopsis.

There are several key questions that arise from the above list of pitfalls. They have important implications for the design, development, and roll out of new systems.

C. What should district leaders know about potential pitfalls in data warehouse design?

The Teacher Incentive Fund (TIF) grantees and others contemplating compensation reform represent a diverse and varied set of districts, states, and charter schools. While these projects differ on many significant factors such as performance metrics and incentive structures, they all present significant implications in the area of information technology and more specifically, data warehousing. The grant applications under the TIF require either significant development of existing data systems or new systems development, especially in the area of data warehousing. Many of the proposals envision a comprehensive incentive framework that will require an integrated view of data from disparate data systems, including student assessment data on statewide exams, enrollment data, teacher certification data, student demographics, registration data, and professional development data of district staff.

In addition to the data requirements of a comprehensive pay-for-performance plan listed above, the metrics that might be used to ascertain individual performance imply a value-added approach. Crucial to the ability to perform value-added analysis is the capability to track students over time. Most school districts, in response to *NCLB* legislation, have implemented student ID numbers. However, deploying a unique identifier for students is not the panacea that some suggest. In fact, unique identifiers will raise more questions about how to track students, teachers, and administrators. For example, one challenge that many TIF grantees face is implementing a process for tracking mobile students across districts. School districts need to be able to retrieve previously assigned numbers for returning students. This can be a particular challenge for urban school districts with high student mobility s(i.e., the workflow requires a larger percentage of human capital). The student ID number should also be consistent throughout all data applications. Many times a student information system, student assessment system, and other applications (e.g., food service) will all use separate identifiers.

Table 3. Data Warehouse Pitfalls[23]

Developing multiple point solutions (aka stove piping) – Using data marts to provide access to each source system can inhibit integration across systems.

Solution: *Focus on enterprise-wide data definitions and use standardized dimensions that can be used across the whole organization. A conformed dimension table used with different fact*

Unconstrained scope – Trying to specify users' needs up front of design is very difficult because 1) not all needs are well known or understood, and 2) needs may change during the time it takes to implement.

Solutions: *Divide development into stages that have clearly defined scope and keep reviewing progress to ensure that development is not creating stove pipes where integration is sacrificed.*

Lack of scalable infrastructure – Initial performance will likely be good, until more data are added and users increase traffic. Infrastructure has three areas: hardware, software, and network connectivity (bandwidth).

Solutions: *Do not plan on sharing hardware platforms between operational and relational databases. Communicate with experts and colleagues who have built similar systems. Identify ways to benchmark performance. Do not choose warehouse components before you understand business requirements.*

Unmanageable administration – Administration tasks include loading, backup, and archival processes, as well as managing new data flows, cleansing, and data quality maintenance. Security, documentation, and training/supporting users are also administrative tasks.

Solutions: *Assign a data warehouse administrator who has full control over the data and data definitions. Assign adequate resources to support this position. Do not allow data quality to degrade, as it will diminish utility and defeat the purpose of having a data warehouse in the first place.*

Resource contention – The data warehouse may get leveraged to address operational tasks that are not well supported elsewhere. This can lead to a degradation of the data warehouse mission and function as well as competition for IT funding and development resources.

Solutions: *Develop an operational data store to support operational needs. Segregate operational and decision support environments. Identify boundaries between operational and decision support functions.*

Inadequate change management – With Online Analytical Processing (OLAP) software, the more adventurous may learn to interrogate the data themselves, following a train of thought that was not anticipated, let alone built into the system. This paradigm shift towards self-service reporting and quantitative analysis can lead to significant process and cultural changes, and may encounter resistance from both IT staff and business users.

Solutions: *Engage senior management and users in design and development process.*

Selecting the wrong reporting and analysis tools – Balancing the reporting needs of power users (e.g., data analysts who create graphs, tables, and reports) with casual users (e.g., data consumers who read and use reports) can be tricky. While power users needs are important, they should not overshadow casual users' needs. Also, software changes often, and selecting a tool too soon can limit access to functionality that is developed in the near future.

Solution: *Focus on data structures and quality because problems there make reporting and analytic functionality moot.*

A similar issue arises with the assignment of teacher identifier numbers. As in the student ID number, the teacher ID number should be consistent throughout all of the data applications within the district. This problem is particularly acute when attempting to link teachers to students. Many times, the teacher ID is a product of the district human resource system and is completely disconnected from the student information system where student-to-teacher links are typically made. The lack of quality data can be compounded by team-teaching, student and teacher mobility, and place holder assignments in cases in which the teacher is unknown at the time of assignment. In some cases, school districts have been able to work around this problem by using alternative systems that teachers must use that establish the required link. Examples include electronic grade books and attendance reporting systems.

The student ID and teacher ID problems point to the larger issue of semantic differences in similarly named attributes.[iv] These differences can range from business definitions (e.g., mandatory versus optional student fees) to data domain differences (e.g., a current student address versus a graduate address). The implementation of a searchable, easily maintainable data dictionary would address many of these issues (see Data Dictionary Overview). Development of a data dictionary will also facilitate formulation review of definitions of concepts that currently may be only vaguely defined. An example of this would be defining the scope of activities that constitute professional development. If one were interested in the effectiveness of different professional development options, a dictionary application and review process can help to make the development of a model more transparent.

Integration of information from several disparate systems requires a normalized database that provides an integrated, consistent, and non-redundant representation of all the data systems. A normalized database will provide a flexible, scaleable platform that will handle electronic transformation and loading (ETL) of external data feeds as well as provide a workspace for data cleaning. The relational data store will also implicitly represent the data rules of the district. An additional advantage of constructing a normalized operational data store is that insertions and updates are highly efficient. The normalized database will serve as a research data warehouse and will be the source for generation of data marts that will support generalized, high-demand reporting (such as schools that make adequate yearly progress) and specialized value-added analysis. There are a few rules for database normalization; each rule is called a normal form. If the first rule is observed, the database is said to be in *First Normal Form*. If the first three rules are observed, the database is considered to be in *Third Normal Form*. Although it is possible to normalize

[iv] The term *attribute* is used as a technical term to describe what might also be called variables. One of the units of analysis in the database example provided here is a person. A person can have many attributes — some with similar names, but different uses in practice. Both student ID and teacher ID are attributes. They are related to roles a person takes in the system.

a database to fourth and fifth normal form, database normalization beyond third normal form is rarely considered.

First normal form requires that there can be no multi-valued attributes and no repeating groups. The table should have a column or combination of columns that will allow unique identification of each row.

Consider the following SchoolTeacher table:

School name	Teacher name
Walter Johnson	Jones, Smith, Einstein
Montgomery Blair	Heisenberg, Maxwell, Dirac
BCC	Teller, Oppenheimer, Smith

In this table, the Teacher Name field is a multi-valued attribute. Now consider this SchoolTeacher table:

School name	Teacher1	Teacher2	Teacher3
Walter Johnson	Jones	Smith	Einstein
Montgomery Blair	Heisenberg	Maxwell	Dirac
BCC	Teller	Oppenheimer	Smith

The Teacher1, Teacher2, Teacher3 table represents a repeating group of similar data. This design will only accommodate three teachers per school.

Now consider this final example:

School name	Teacher name
Walter Johnson	Jones
Walter Johnson	Smith
Walter Johnson	Einstein
Montgomery Blair	Heisenberg
Montgomery Blair	Maxwell
Montgomery Blair	Dirac
BCC	Teller
BCC	Oppenheimer
BCC	Smith

This design is in first normal form since there are no multi-valued fields, there are no repeating groups, and the combination of School Name and Teacher Name as a primary key uniquely identifies each row.

Second normal form requires that non-key fields must be functionally dependent on the entire primary key. Consider the following StudentCourses table:

Student ID	Course ID	Student name	Course title	Grade
12345	1300	Bob	Physics	A
12345	3100	Bob	Calculus	B

In this example, the combination of StudentID and CourseID is the primary key. However, StudentName has no dependency on CourseID. The CourseTitle has no dependency on StudentID. Grade is the only field that is functionally dependent on the CourseID and StudentID. To satisfy the second normal form rules, the data should be split into three separate tables:

Students Table

Student ID	Student name
12345	Bob

Courses Table

Course ID	Course name
1300	Physics
3100	Calculus

Student Courses Table

Student ID	Course ID	Grade
12345	1300	A
12345	3100	B

Third normal form prohibits transitive dependencies. A transitive dependency exists when any non-key attribute in a table is dependent on another non-key attribute. Consider the following CourseSections table:

Course Sections

Course ID	Section	Teacher ID	Teacher name
3100	1	23	Maxwell
1300	1	25	Newton

In this example, CourseID and Section uniquely identify TeacherID. However, TeacherName is dependent on TeacherID and has no dependency on CourseID and Section. This is a transitive dependency and a violation of third normal form. The data are properly stored in two tables:

Teacher Table

Teacher ID	Teacher name
23	Maxwell
25	Newton

Course Sections

Course ID	Section	Teacher ID
3100	1	23
1300	1	25

While it is theoretically desirable to have all tables adhere to third normal form, it is not always practical. Consider a table that contains teacher home addresses:

Teacher Address

Teacher ID	Street address	City	State	Postal code
1	1025 State St.	Madison	WI	53706

TeacherID is the primary key, and the table is in second normal form. TeacherID will uniquely identify the street address, city, state, and postal code. The postal code can also be uniquely determined from the street address, city, and state. This is a transitive dependency and a violation of third normal form. To correct this, you must create a separate table for the postal codes to avoid update anomalies. In *real world* design, it often is more practical to apply third normal form rules to data that change frequently, rather than add the overhead of additional tables for data that change infrequently, if at all.

Normalization, as described above, is a systematic process of organizing data in a database. The process includes creating tables and establishing relationships between those tables according to rules designed both to protect the data and make the database flexible by eliminating two factors: redundancy and inconsistent dependency.

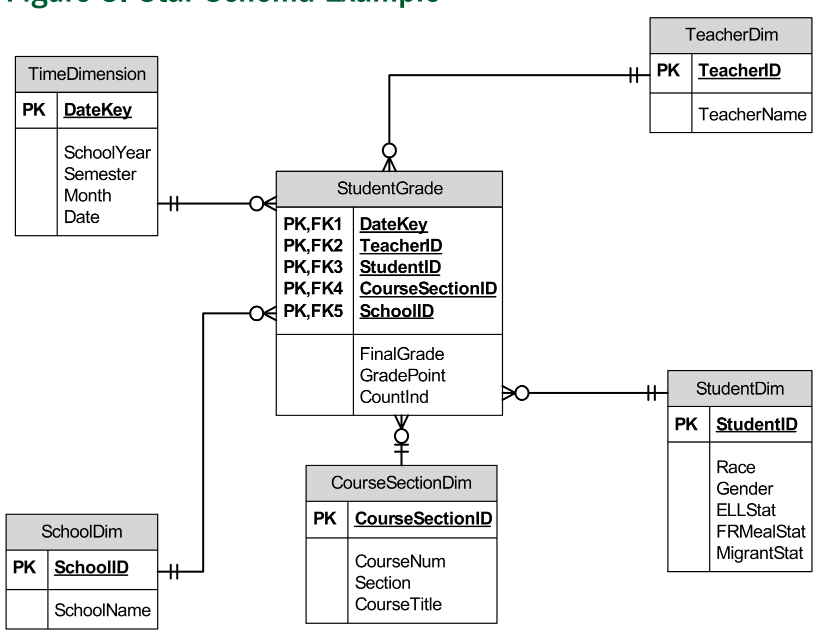
Redundant data waste disk space and create maintenance problems. If data that exist in more than one place change, all locations must be updated in exactly the same way. For example, an update to an individual teacher's *highest degree achieved* is much easier to accomplish if that piece of information is stored only in the teacher table and not in the teacher assignment, professional development, and teacher portfolio systems as well.

Inconsistent dependencies can make data difficult or impossible to access; the path to find a piece of data may be missing or broken. For example, it is intuitive to look in the *School* table to find the name of a particular school, but it makes no sense to search the *School* table to find the name of a teacher working at that school. A teacher's name is related to, or dependent on, the teacher and should be stored in the *Teacher* table.

The data in a database normalized to third normal form are optimized for very quick insertion and update of discreet data items, which supports loading new data into the warehouse and data cleaning. Theoretically, a data set meeting the requirements of virtually any reporting or analysis project can be generated from a normalized database. In practice, smaller subsets of the data warehouse are constructed using a design concept called a dimensional model. A dimensional model has the advantage of being easy to use, and retrieval performance is generally superior to a normalized database.

Most dimensional models implement a design pattern known as a Star Schema. Consider the following design:

Figure 3. Star Schema Example

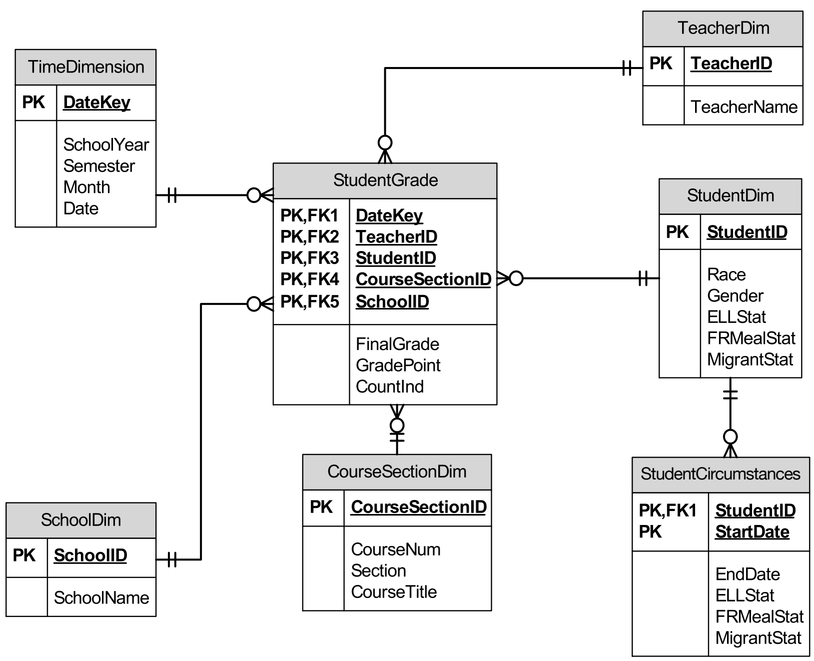


This design has a fact table called StudentGrade and several related dimension tables. This design supports rapid retrieval of student grade information. The CountInd column is present to support counting of rows.

The dimension tables support the *slicing* of information by school year, school, teacher, or any number of student attributes. This design does not support efficient updating or insertion of new information.

In the typical Star Schema design, the fact table (StudentGrade) contains information concerning behavior (in this case, the grade received). The dimension tables contain information concerning circumstances. For example, the StudentDim table provides information concerning race, gender, ELL status, Free/Reduced-price Lunch status. While some attributes will remain fixed (e.g., gender and race), other attributes will change over time. The ability to track changes in circumstances will facilitate value-added analysis. Tracking changes in circumstances requires that those attributes that can change over time be stored in separate, related tables.

Figure 4. Star Schema with Change In Circumstances



This design will optimally support general reporting and value-added analysis. It will allow investigation of student behavior changes and should be considered as the preferred data mart design in support of the incentive framework proposed in the TIF grant applications.

The information technology infrastructure required to support the teacher incentive programs will require development of several data systems. A normalized data warehouse system that will support electronic loading and data cleaning from disparate source systems will be required. This normalized database will serve as the source for data marts implemented in a dimensional model. The dimensions will need to support the concept of changing circumstances. Additionally, an electronic data dictionary that documents the source, definition, and semantics of all terms will be required.

Conclusion

In general, we have found IT problems the most consistently difficult to discover, diagnose, and correct. Many districts do not routinely include research, assessment, human resources, or technology directors in cabinet-level meetings. Many district (and state education agency) staff members have strong assumptions about the quality of district data, based on years of successful compliance with state and federal reporting demands. The challenge for those taking on these new tasks will be to constantly question the quality of the data being used to make these high-stakes decisions. More important, the district leaders associated with the collection, management, and analysis of these data need to be at the table to help address the problems at their source and all along the process.

End Notes

[1] Wayman, J.C., Midgley, S., & Stringfield, S. (2005). *Collaborative teams to support data-based decision making and instructional improvement.* Paper presented at the 2005 annual meeting of the American Educational Research Association, Montreal, Canada, Retrieved January 9, 2007, from http://www.csos.jhu.edu/beta/datause/papers/waymancollaera.pdf

[2] http://www.sifinfo.org

[3] This notion of tracking productivity of a educational system is based on the concept of the *education production function*. For more explanation of this function (as well as the role of incentives to improve one's own human capital) see, for example, Wilson, K. (January 2001) The determinants of educational attainment: Modeling and estimating the human capital model and education production functions. *Southern Economic Journal, 67*(3), 518-551 or Hanushek, E.A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature,* 24, 1141-1177.

[4] English, L. (2002). The essentials of information quality management. *DM Review*, 12(9), 36-44.

See also Inmon, W.H. (2006). *Data warehouse 2.0*. Castle Rock, CO: Inmon Data Systems. May 11, 2007; Karacsony, K. (2006, February 24). Proactive data quality. *DM Review*; Mullen, N. (2003, January). Information for innovation: Less bitter: Improving the quality of your data. DM Review.

[5] English, L. (2002). The essentials of information quality management. *DM Review*, 12(9) 36-44.

[6] See http://www.fairtest.org/facts/Limits%20of%20Tests.html for a discussion of the problems associated with using standardized tests for diagnostic purposes.

[7] See, for example, Inmon, W.H. (2006). *Data warehouse 2.0*. Castle Rock, CO: Inmon Data Systems. May 11, 2007.

[8] Robertson, G. (2007). Dataless hybrid approach solves many universal data integration, sharing and interoperability problems. *DM Review*. Retrieved January 2, 2008, from http://www.dmreview.com/article_sub.cfm?articleId=1077688-1.html

[9] Kesner, R.M. (2005). *Decision support systems (DSS) deployment strategies* (Working Paper). Boston: The Center for Information System Deployment, Northeastern University.

[10] See, for example, Bhatt, G.D., & Zaveri, J. (2002, January). The enabling role of decision support systems in organizational learning. *Decision Support Systems, 32*(3), 297-309; Hall, M. (2002, July 1). Decision-support systems. *Computerworld, 36*(27), 31; and Kalay, P., & Chen, D. (2002, Summer). Integrating a decision support system into a school: The effects on student functioning. *Journal of Research on Technology in Education, 34*(4), 435-453.

[11] Wayman, J.C., & Stringfield, S. (2006). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education, 112*(4), 549-571.

[12] http://specification.sifinfo.org/Implementation/2.0/StudentInformationSystemsWorkingGroup.html#StudentInformationSystemsWorkingGroup; Karbowski, S. (2005). Student information systems: Data-driven decisions in K-12 schools. *Media and Methods, 41*(4), 8; Nobles, K. (2004). Student data information systems: Expanding what you already have. *Media and Methods, 41*(1), 18; Schools Interoperability Framework Association. (2007). *Schools Interoperability Framework™ implementation specification 2.0r1.* Retreived August 15, 2007, from http://specification.sifinfo.org/Implementation/2.0r1/

[13] Petrides,L.A. (2002). Organizational learning and the case for knowledge-based systems. *New Directions for Institutional Research, 11*(3), 69-84.

[14] See Hamblin, E.M. (2001). *The HelpWorks™ return on investment: Social and economic benefits of implementing a rules-based benefit outreach strategy.* (White Paper). Chicago, IL, Peter Martin Associates.

[15] Hamblin, E.M. (2001). *The HelpWorks™ return on investment: Social and economic benefits of implementing a rules-based benefit outreach strategy.* (White Paper). Chicago, IL, Peter Martin Associates.

[16] Cochrane, J.D. (2001). Delivering patient care on the Internet. Retrieved January 4, 2003, from http://www.epicsystems.com/news/pdf/ht_sys_09_01.pdf

[17] http://www.epicsystems.com/

[18] Davenport, T.H., De Long, D.W., & Beers, M.C. (1998). Successful knowledge management projects. *Sloan Management Review, 39*(2), 43-58.

[19] Davenport, T.H., De Long, D.W., & Beers, M.C. (1998). Successful knowledge management projects. *Sloan Management Review, 39*(2), 43-58.

[20] There are a number of online resources on that describe the creation and use of online educational content and applying standards. See the following examples which also provide links to many other related resources: http://www.eduworks.com/LOTT/tutorial/index.html & http://www.learnativity.com/standresources.html

[21] In fact, efforts to isolate the return on investment of information technology are still not being widely done. See, for example, McKnight, W. (2004, May). Building business intelligence: BI business value: Does it come from the program or the projects? *DM Review*.

[22] Davenport, T.H., De Long, D.W., & Beers, M.C. (1998). Successful knowledge management projects. *Sloan Management Review*, 39(2), 43-58.

[23] For more details see http://www.orsoc.org.uk/about/topic/projects/elwood/Pitfalls.htm#top